

Formative Assessment of Writing (FAoW): A Confirmatory Factor Structure Study

Elaheh Tavakoli ¹, Mohammad Reza Amirian ^{*,1}, Tony Burner ²,
Mohammad Davoudi ¹, Saeed Ghaniabadi ¹

¹ Hakim Sabzevari University, Sabzevar, Iran

² University of South-Eastern Norway, Norway

ARTICLE HISTORY

Received: 24 March 2019

Revised: 21 May 2019

Accepted: 09 July 2019

KEYWORDS

Formative Assessment of Writing,
Factor Solution,
Formative Assessment,
Confirmatory Factor Analysis (CFA)

Abstract: This validation study was undertaken to evaluate the construct of Formative Assessment of Writing (FAoW) operationalized by an instrument with 50 Likert scale items. To identify the EFL learners' experiences of FAoW practices, the instrument was first piloted with three EFL learners, and subsequently administered on a sample of 255 EFL learners selected based on purposive sampling. A five-factor solution with five latent variables (i.e. clarifying criteria, evidence on students' current learning, feedback to move learners forward, peer-assessment and autonomy) was evaluated through Confirmatory Factor Analysis (CFA) with AMOS 22. Model fit showed that the five-factor structure of FAoW could only be supported in terms of absolute and parsimony fit indices. The model with three factors (i.e. clarifying criteria, peer-assessment and feedback) in two stages of pre- and while-writing, however, provided higher discriminant validity in addition to absolute and parsimony fit indices. In other words, FAoW was not found to be practiced within its full potential with five components in the context of this study. A conceptual model was developed based on the findings and the literature to show pedagogical application of FAoW and how it can be practiced in line with Black and Wiliam's (2009).

1. INTRODUCTION

Since Formative Assessment (FA) was introduced to the field of education in the late 1990s by the Assessment Reform Group in the UK (e.g. Black & Wiliam 1998), many scholars, particularly in Europe and the USA, tried to investigate its theoretical base and practice. Most importantly and more closely related to this study, Black and Wiliam (2006; 2009) tried to provide a unifying theoretical framework for FA practices after interviews with teachers who developed FA and observation of the changes that occurred in their classrooms.

In second language (L2) writing, however, FA has been underexplored and much of the available research has focused on summative assessment, peer assessment or the effectiveness

CONTACT: M. R. Amirian ✉ smr.amirian@hsu.ac.ir 📧 Department of English Languages and Literature, University of Hakim Sabzevari, Sabzevar, Iran

of teachers' feedback (Burner, 2015; Lee, 2003; 2011). Formative Assessment of Writing (FAoW) is a prospective and aims to improve learning and fill the gap between students' current and potential state of development. It is a construct which has not been adequately defined, operationalized and validated so far. This study is a response to Johnson and Riazi (2017), who referred to the lack of local validation efforts for ensuring that the writing instruments are compatible with the unique learning outcomes, students, and context of the program. Tavakoli, Amirian, Burner, Davoudi and Ghaniabadi (2018) developed a FAoW instrument (Appendix I) which consisted of a comprehensive list of FA practices in writing classrooms based on Black and Wiliam's (2009) FA framework and Hattie and Timperley's (2007) feedback model. This study is factor structure of that instrument and part of a PhD project to investigate FAoW from both teachers' and students' perspective. In the project, two parallel versions of FAoW instrument were developed: EFL students' experiences of their teachers' FAoW practices and EFL teachers' perspective about their own practice of FAoW. Our earlier article Tavakoli, et al., (2018) pertained to the theoretical foundation and the development of FAoW instruments. This research is an attempt to validate the students' version through CFA. In this study, the words item, experience and practice are used interchangeably as every item in FAoW instrument is a teacher's FA practice or classroom activity which the students reported the frequency of their experience.

1.2. Review of the literature on FAoW

The literature on FAoW has highlighted some studies (e.g. Burner, 2015; 2016; Lee, 2007; 2011; Lee & Coniam, 2013; Mak & Lee, 2014; Naghdipour, 2016; 2017; Saliu-Abdulahi, 2017; Saliu-Abdulahi, Hellekjær, & Hertzberg, 2017; Tavakoli, et al., 2018; Wingate 2010). The construct of FA in general has been described and conceptualized in various ways (Bennett, 2011); different scholars have developed different writing assessment instruments which could be used formatively. However, "there is no one definition of formative assessment of writing" (Burner, 2016, p. 4).

In the current research, we probed into FAoW considering FA model in general and assessment practices and feedback on the students' writing assignments in particular. The construct of FAoW has been operationalized in some studies which are worth citing here. In line with the ten principles of FA (aka Assessment for Learning) by Assessment Reform Group (2002), Lee and Coniam (2013) described the implementation FA in writing in terms of three phases: 1. Teachers' cooperative planning of the teaching resources and feedback forms; 2. instruction based on the teaching-learning cycle (setting the context, modeling and deconstruction of texts, joint construction, and independent construction) and 3. actual writing assessment phase using the same criteria at the instructional stage. In another study, Mak and Lee (2014) examined EFL teachers' implementation of FAoW in six classrooms over a course of one year through classroom observations and interviews with administrators and teachers. The schools adopted a FA plan with three phases of the writing process_ pre, during and post-writing stages. In pre-writing stage, teachers familiarized the students with the assessment criteria and set their goals. In the during-writing stage, the students benefitted from their peers' and the teacher's feedback and used their focused and coded corrective feedback. The feedback corresponded with the assessment criteria which had been established in the pre-writing stage. In the post-writing stage, the students recorded the number of errors in their error log and reflected on their progress. The act of reflection also involved students in thinking critically about their own writing and the feedback they received from both their peers and teachers so that they could make use of the information to feed forward and benefit their future writing. With the three staged research plan, the teachers were consequently able to teach what they assessed and assess what they taught.

In an earlier study by Tavakoli et al. (2018) based on Black and Wiliam's (2009) FA framework and Hattie and Timperley's model of feedback (2007), FAoW was operationalized in an instrument to measure EFL students' experience of FAoW practices as their role along with the teachers' in FA is of crucial significance and, according to many scholars (e.g. Feng, 2007), this role has been overlooked. Brookhart (2001) placed the students in the central role and, in line with Black and Wiliam (1998), considered assessment to be formative only when the information it provides is used for improving students' performance and learning. She explained that the limited research on the role of students in FA is probably because teachers are always considered to plan and administer classroom assessments. This study, as part of a bigger project, is to fill the gap and investigate the construct validity of a FAoW instrument which measures teachers' practice in the view of EFL learners in writing classes.

Assessment of writing has been documented by some researchers in the EFL context (e.g. Elahinia, 2004; Ghoorchaie, Tavakoli, & Nejad Ansari, 2010; Javaherbakhsh, 2010; Mosmery & Barzegar, 2015; Moradan & Hedayati, 2011; Naghdipour, 2016, 2017; Nezakatgoo, 2005; Sadeghi & Rahmati, 2017; Sharifi & Hassaskhah, 2011). Most of these studies on FA in writing classrooms have been experimental case studies on the effect of FA practices when introduced through an intervention or qualitative researches on the existing assessment practices in writing classrooms. For instance, Naghdipour's (2016) interviews with teachers and students showed that FA tools such as collaborative tasks, portfolio writing, and other process- and genre-based strategies were absent in the EFL writing classrooms. In another attempt (Naghdipour, 2017), FA was incorporated into a university EFL writing course and the data on students' beliefs and attitudes were collected through semi-structured interviews at the end of the semester and pre- and post-study attitude questionnaires (developed mainly in line with Lee, 2011). FAoW intervention was a three-session modular instruction to teach writing based on five FA strategies outlined by Black and Wiliam (2009). First, pre-writing stage of instructional tasks which made students write based on model essays, brainstorming and pooling of ideas, (see Naghdipour & Koç, 2015, for an overview). The second draft for each task was written in response to the peer-assessment and the third draft was revisions after the teacher assessment. FA intervention revealed an improvement in various aspects of participants' writing and development of their positive attitudes towards writing as well as FA.

There is a consensus in many studies on the beneficial effect that alternative forms of assessment. However, when implementing various forms of formative assessment is explored for writing classrooms, the existing researches fail to account theoretical models and operationalized set of FAoW practices for EFL context.

Operationalization of FAoW construct and the development of an instrument to measure FAoW was the focus of another study by Tavakoli et al. (2018). The information on the development of FAoW instrument is crucial to this research as this study aims at the construct validity of that instrument through a confirmatory approach and model fit. The instrument (Appendix I) was developed through an intuitive approach with 50 items in an earlier study. The items were classified under 5 factors (colored differently in [Table 1](#)) through a focus group interview with three EFL experts in the domain of assessing writing. The experts agreed on the five dimensions underlying the items in the instrument and indicated FAoW to be multidimensional. [Table 1](#) illustrates the items under the five FA factors and the three writing stages.

Table 1. FAoW framework, item and construct matching by experts (Adapted from Authors, 2018)

	Where the learner is going? Pre-writing (feed up)	Where the learner is right now? Writing (feedback)	How to get there? Post-writing (feed forward)
Teacher	Items 1, 2, 3, 4, 5, 6, 7, 8, 12 Clarifying criteria	Items 14, 15, 18, 20, 22, 23, 29, 30, 31, 36, 40, 43, 48 Evidence on students' current learning	Items 32, 33, 37, 39, 41, 44, 45, 47, 50 Feedback to move learners forward
Peer	Items 9, 10 Clarifying criteria	Items 16, 17, 25, 26, 28 Peer-assessment	
Learner	Items 11, 13 Clarifying criteria	Items 19, 21, 24, 27, 34, 35, 38, 42, 46, 49 Autonomy	

The notions of ‘feed up, feedback and feed forward’ corresponded with the main function of FA to “reduce discrepancies between current understandings and performance and a goal” (Hattie & Timperley’s, 2007, p. 86). As Table 1 illustrates, thirteen items tapped pre-writing stage activities such as model-writing, pre-writing planning, setting writing goals, organizing and developing writing ideas, free writing, and clarifying assessment criteria. These writing activities related to ‘feed up’, defined as ‘the goals one lays down to achieve’ (Hattie & Timperley, 2007, p.86) and corresponded with clarifying criteria in Black and Wiliam’s (2009) FA framework. Students set attainable goals so that they understand what they are working towards in the ‘feed up’ stage (i.e., where they are going).

‘Feedback’/ while writing stage guided the second set of items in FAoW instrument and specified assessing the progress that was being made towards the goal. The items included writing practices such as process-writing/ multiple drafting, writing feedback on progress, peer-writing feedback, writing error log, computer feedback, autonomous writing revision, writing reflection and self-assessment. Thirty items were placed under this construct and tapped a variety of feedback (e.g. graded, focused, indirect, direct and descriptive) from various sources (e.g. peers, teachers and the learners). This stage of writing corresponded with ‘where the learner is right now’ principle of FA and implied the learners’ prior progress and current state of learning.

Items for assessing students’ performance at post writing stage encompassed those practices which could lead students for their future improvement such as reflection for future progress, teacher-oriented feedback and portfolio assessment. They corresponded with ‘feed forward’ in the feedback model and, as Mak and Lee (2014) confirmed, covered writing practices which gave students a direction of what they were to achieve in the future, a blueprint of where they were going in the future.

To date, factor structuring of FAoW based on a unified theoretical framework has received scant attention in the literature. This indicates the need to factor structure the construct of FAoW through CFA. Hence, this study aimed at both theory verification and modification in an EFL context and attempted to answer the following research questions:

1. Does the five-factor FAoW model fit the data collected from EFL students of writing courses?
2. What is a model describing EFL teachers’ practice of FAoW in the view of EFL students?

2. METHOD

2.1. Participants

Since the researchers aimed at assessment of the students' writing assignments at discourse level, the criterion for selection of the participants for both research questions was their prior experience of writing tasks at the level of paragraphs and essays. This made the researchers select junior and senior university undergraduates and upper-intermediate or advanced level language school students in Iran. Based on the prior experience of writing assessment criterion, sampling of participants for both the interviews and the quantitative data collection was purposive.

For piloting FAoW instrument, three EFL students (two from language schools and one from a university) were selected. All the interviewees were female and the researchers resorted to the same criterion of having writing assessment experience in their selection.

For responding to both research questions, a purposive sample of 315 Iranian EFL students was selected from three non-state language schools and five universities. Of the initial cohort, 255 respondents had more reliably and completely filled in the instrument (response rate of 85%). The participants' age ranged from 13 to 48 ($M=22$). Overall, sixty-seven of them were males and one hundred eighty-eight of the participants were female (See Table 2).

The selection of participants was based on the criterion of prior experience of writing assessment. Participants from universities were selected from senior and junior undergraduate students of English since the English curriculum in Iran requires students to pass three mandatory writing courses (Advanced Writing, Essay Writing and Paragraph Writing) in the first two years. All the participants had finished writing courses/lessons at discourse level and had experienced assessment of their essay writing tasks prior to completion of FAoW instrument. Furthermore, selection of students from language schools was based on the level of English textbooks they covered at the time of data collection (upperintermediate and advanced based on CEFR[†]) and the agreement of researchers that the books and the school curriculums included writing tasks at the level of discourse.

As shown in Table 2, 66.2% of the participants were learning English writing through the university undergraduate curriculum (Teaching, Literature or Translation of English) and 33.7% of them through private language schools.

Table 2. *Participants Demographic Information by number (%)*

Context of education	N (%)	Gender		Education
		Male	female	
University*	169 (66.2)	28(16.6)	141 (83.4)	Teaching English= 47(27.8) English Literature= 27(16) English Translation=95(56.2)
Private Language School**	86 (33.7)	39 (45.3)	47 (54.7)	Highschool= 39(45.3) Diploma= 20 (23.2) Bachelor= 18(20.9) Masters=18 (20.9)
Total	255 (99.9)	67 (61.9)	188 (138.1)	

*Third and fourth year bachelor students of English

** Upperintermediate and advanced level students

[†] in CEFR (Common European Framework Reference), B2, C1 and C2 define upper-intermediate an advanced level of English

2.2. Instrument tool

FAoW instrument taps at students' experiences of FA practices in writing classrooms and their attitudes towards the helpfulness of each practice. It had initially been developed by Tavakoli et al. (2018) with five underlying constructs based on Black and Wiliam's (2009) Formative Assessment (FA) and Hattie and Timperley's (2007) feedback model. In line with the intuitive approach of scale construction (Hase & Goldberg, 1967), a comprehensive review of the literature was undertaken, and 50 Likert scale items were devised. Three experts in the field of writing assessment intuitively classified the items based on the five components of FA (clarifying criteria, evidence on students' learning, feedback to move learners forward, peer assessment and autonomy) and in three stages ("Where the learner is going/Pre-writing, "Where the learner is right now/Writing and "How to get there/ Post-writing"). The estimates for the current study were only derived from the students' responses to the four-point Likert scales under experience.

FAoW questionnaire which was developed by Tavakoli et al. (2018) and used in this survey consisted of two sections. Section I solicited details on the participants' demographics such as age, gender, writing and assessment experience and their highest level of academic qualification. Sections II was the items which sought to determine students' experience of and attitude towards FAoW. It was rated by EFL learners using a four-point Likert type scale for experience (ranged from 1 to 4) on the left and the scale of attitude (ranged from 1 to 5) on the left side of each item. In the experience scale, 1 was a practice that had never been experienced, 2 was rarely, 3 for often and 4 as a FAoW practice that had always been experienced by students in their writing classrooms. The attitude scale measured students' attitude from 1(very unhelpful) to 5 (very helpful).

While the development of the instrument and qualitative operationalization of its construct was the focus of the authors' earlier study (Tavakoli, et al., 2018), this study aimed at its quantitative construct validation. Here FAoW was piloted by three EFL learners before the large-scale administration and CFA in order to identify if the language of the instrument was comprehensible to EFL learners. Three EFL learners were interviewed separately; they were asked to read each item and explain or exemplify their understanding of each FAoW practice either in their first (Persian) or foreign language (English).

2.3. Data collection procedures

The development of the instrument and piloting it with the EFL learners took place in the first semester of 2015-2016 academic year. Afterwards, the interview with the three EFL learners were independently conducted. Each interview lasted 70 minutes on average, was audiotaped and transcribed verbatim for further analysis.

After the interviews, paper FAoW instrument (Appendix I) was distributed among EFL learners in both language schools and universities with the attendance of the first researcher to provide assistance in case required. There was no time restriction to complete the instruments, but filling out the instrument took approximately 35 to 45 minutes.

2.4. Data analysis

Using SPSS 19 Cronbach's index of internal consistency was estimated for internal consistency of FAoW instrument. To respond to the research questions, that is, to factor structure and verify Black and Wiliam's (2009) model and evaluate the model currently employed by teachers, CFA was run on the students' survey data using Analysis of a Moment Structures 22 (AMOS). To construct validate the instrument through CFA, the missing data was handled first. As 8.5% of the data (above 5%) were missing and MCAR test revealed nonrandom missing (Little's MCAR Test: Chi-Square= 3430.96, DF=3038, Sig. = .000), series mean method could not be used to handle missing data (Tabachnick & Fidell, 2007). Hence, single imputation had to be used to

screen the missing data. The data were also checked for kurtosis, skewness, normality and outliers. Although the distribution of data was found to be normal for all variables with the skewness of all experiences within the acceptable range of +3 and -3, multivariate normality and linearity test revealed 28 outliers/cases ($p < .05$), which were removed from the subsequent analysis.

From several types of parameters which are commonly reported to indicate goodness of fit for measurement models, for evaluating the FAoW model in this study we report one index for every of the three broad categories. Root Mean Square Error of Approximation (RMSEA) is reported for absolute fit which calculates the standardized residuals resulting from fitting FAoW model to the data. Comparative Fit Index (CFI) for relative fit is reported as it adjusts for the issues of sample size inherent in the chi-squared test of model fit and the normed fit index. CFI analyzes the model fit by examining the discrepancy between the data and the hypothesized model and indicates better fit when it is closer to 1. Finally, standardized root mean square residual (SRMR) was ultimately used as a parsimony fit index in this research. This is in line with Brown (2015) who advise researchers to consider and report at least one index from each category when evaluating the fit of their models. Because chi-square is the basis for most other fit indices, it is routinely reported in all researches as an original fit index for Absolute fit (Brown, 2015), it is reported in addition to the three indices.

While absolute fit indices do not use an alternative model as a base for comparison, relative fit indices compare a chi-square for the model tested (FAoW with five constructs) to “baseline”, “independence” model (Aka null model) with no latent variables in which all measured variables are uncorrelated. Parsimony-corrected fit indices are relative fit indices that are adjustments to most of the formerly-mentioned fit indices.

3. RESULT

3.1. The pilot study

Qualitative analysis of the interviews with the three EFL learners confirmed their understanding of the FAoW practices underlying the five constructs (Table 1) particularly with the help of definitions or examples which were provided for every item. In each interview, special attention was paid to the clarity of key terms which corresponded with the constructs. Although the language of the instrument was English, technical terms had been defined, exemplified or translated into the participants' first language. Their verbal reports while reading each item and their admission at the end of interviews showed that despite the apparently confusing technical terms such as ‘assessment criteria’, ‘outline or mind map’, ‘free-writing’, ‘descriptive feedback’, ‘error log’, ‘portfolio’ and ‘qualitative feedback’, further definitions and exemplifications in the instrument extensively added to their understanding of the FAoW practices.

3.2. Descriptive statistics and the reliability of FAoW instrument

Cronbach's index of internal consistency revealed an alpha value of .91 (Table 3), which suggests a high internal consistency for the instrument. In addition to the reliability for the sum scale, Cronbach Alphas were also computed for the five factors of FAoW, i.e. Clarifying criteria, Evidence on students' current learning, Feedback to move learners forward, Peer-assessment and Autonomy, which, except for peer-assessment, showed an acceptable internal reliability (Values above .7 are considered acceptable, though values above .8 are preferable, Pallant, 2007) (Table 3).

Table 3. Reliability and descriptive statistics for FAoW instrument (with 50 items in five factors)

	Total Items		Five FAoW Factors			
		Clarifying criteria	Current learning evidence	Feed forward	Peer-assessment	Autonomy
Cronbach's Alpha	.91	.75	.72	.77	.60	.75
N of items	50	13	13	9	5	10
N of participants	255	255	255	255	255	255
Range	107.09	28.61	29	25	14	25
Minimum	70	18	17	10	5	10
Maximum	177	46.61	46	35	19	35
Mean	113	32.57	27.98	20.42	11.31	21.14
SD	19.48	6.01	5.28	4.82	3.10	5.13

Descriptive statistics in this table also shows the lowest mean (11.31) belonging to peer assessment; however, the decision was made to keep peer-assessment as an underlying section of the FAoW instrument for CFA analysis since, theoretically and based on Black and Wiliam (2009), it is considered as one of the sources of FA and a crucial agent among the three (teacher, peer, learner). Moreover, poor Cronbach alpha in peer assessment is statistically justified as it is attributed mainly to the few number of items (Pallant, 2007).

3.3. Confirmatory factor analysis procedures

Following Black and Wiliam's (2009) framework, a five-factor hypothetical model of FAoW was extracted. The number and nature of latent variables were based on its five components:

- clarifying criteria for success (feed up)
- eliciting evidence of students' current writing ability (feedback)
- providing feedback to move learners forward (feed forward)
- activating students as instructional resources (peer-assessment), and
- activating learners' as the owners of their own learning (Learner autonomy)

The five-factor recursive model was subjected to a confirmatory factor analysis to assess the goodness of model-data fit; the resulted model of FAoW with 50 observed variables/ items and five factors/ unobserved variables is illustrated in the following.

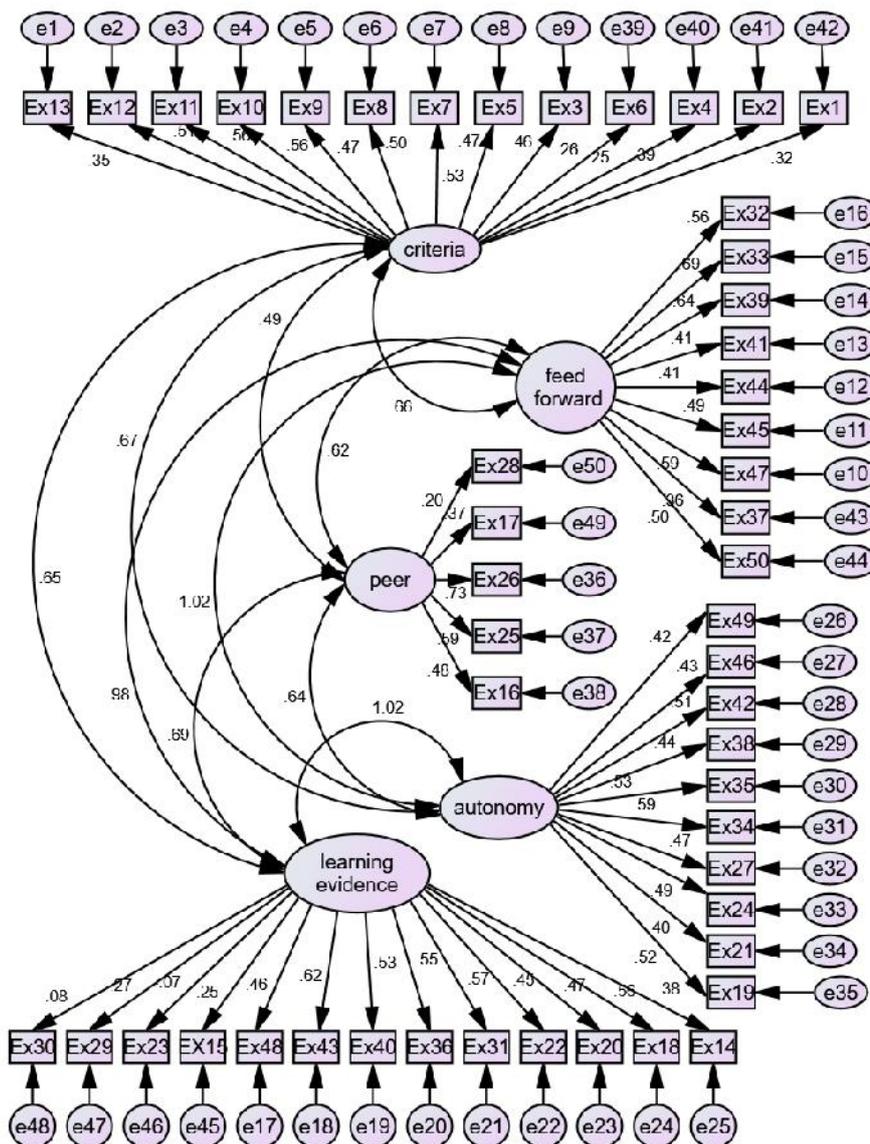


Figure 1. Five factor FAoW model

As Figure 1 shows, CFA model with five latent variables results in three main problems. Firstly, it showed high correlations between three latent variables, i.e. feed forward, autonomy and learning evidence/ feedback ($r = 1.02$, $r = 1.2$, $r = .98$), which is indicative of the three latent variables being only one factor rather than three. Graham, Harris, Fink and MacArthur (2001) explained low factor loadings between latent variables as indicator of a high discriminant validity. The issue of low discriminant validity was handled by merging feedback, feed forward and autonomy into one latent variable (with the label feedback) and trying a three-factor CFA model.

The second problem with five-factor model which encouraged the researchers to try three-factor solution was low factor loadings for eight items with a loading lower than .3. Items 4, 6, 15, 23, 28, 29, 30 and 37 respectively showed loadings of .25, .26, -.25, .07, .20, .27, .08 and .26. All the other loadings between the indicators and the latent factors as well as the covariance among the factors were significant at $p = .001$.

The third problem with this factor structure model was indices of fitness, particularly CFI, which was .70 and lower than the acceptable index (higher than .95, Hu & Bentler, 1999). As Table 4 illustrates, five-factor solution could show only acceptable CMIN, the root mean square error of approximation (RMSEA) and SRMR (respectively 1.85, .056 and .069). Small residuals (RMSEA_.06) indicate a small discrepancy between the observed correlation matrix and the correlation matrix estimated from the model (Hu and Bentler, 1999).

Five factor solution was not shown to have an acceptable comparative fit index nor discriminant validity. Hence, five factor solution of FAoW data statistically showed poor fit with the theoretical models of FA and writing feedback. The aforementioned problems with five-factor solution model made the researchers check three-factor solution through merging items under autonomy, feed forward and feedback and name them “feedback”. Table 4 shows fitness indices and Figure 2 illustrates the model after modification.

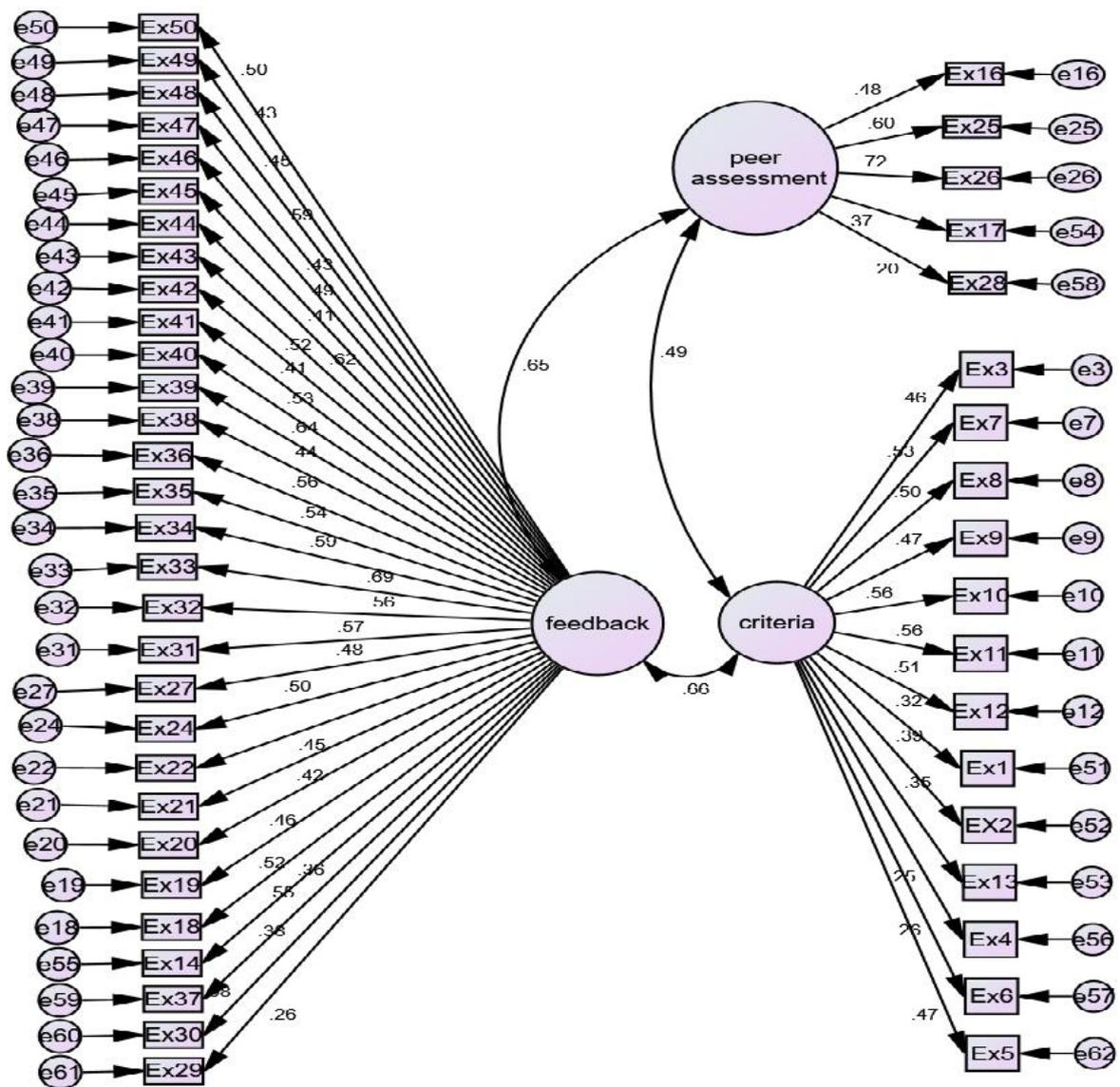


Figure 2. Three-factor model of FAoW

Table 4. Fit indices for five and three factor CFA models of FAoW *

	Absolute fit Indices				Absolute fit Indices	Comparative Fit index	Parsimony Fit index
	Chi-Sq(x^2)	Df	P value	CMIN Chi-Sq(x^2)/Df	RMSEA	CFI	SRMR
Fitting Dataset for five-factor model	2161	1165	.00	1.85	.056	.70	.069
Fitting Dataset for three-factor model	1098	652	.00	1.62	.048	.84	.059
Acceptable threshold levels Hu and Bentler (1999)			P value>.05	1<CMIN<5	close to .06 or below	More than .95	Less than 0.08

*Modified FAoW after removals with factor loadings lower than .4

Although three-factor model resulted in higher discriminant validity with lower correlations latent variables/ factors (.66, .65 and .49), the problem of low standardized factor loadings remained in 8 items (4, 6, 15, 23, 28, 29, 30 and 37). In the modification process, the researchers maintained six items due to their relevance to the construct of FAoW and only removed two items (15 and 23) as they were reverse coded items introduced to the FAoW instrument to eliminate participants' guessing or boredom. More specifically, items 15 and 23 measured teachers' employment of one draft and product writing in contrast to process writing which taps FA. There is the argument in the literature (Brown, 2015) against using reverse coded items in questionnaires as they increase level of measurement error and affect loadings in factor analysis.

Comparison of model fitness indices between the five- and three-factor models of FAoW (Table 4) showed that the latter provided a better fit than the former; particularly in comparative fit index of CFI, which increased to .84, although not within the recommended acceptable range of above .95. The model has improved in discriminant validity as covariance between the three latent variables of feedback, feed up and peer assessment was relatively lower (.65, .69 and .49, Figure 2).

Hu and Bentler's (1999) evaluations criteria were employed for checking goodness fit between the target model and the observed data (see Table 4). Table 4 illustrates model fit indices for both three-factor and five-factor solutions. It reveals that probability or p-value is statistically significant and does not meet the acceptable range for model fit. With the sample size of more than 200, it is difficult to have a non-significant p value since χ^2 statistic is very sensitive to sample size and is not relied upon as a basis for acceptance or rejection. Table 4 also shows that three of the indices (CMIN, RMSEA and SRMR) are within the acceptable range for model fit and confirm the absolute and parsimony fit of both models. Although CFI is lower than the acceptable value in both five and three-factor models (.70 and .84, respectively), three-factor model revealed a better fit in terms of CFI/ comparative fit index.

In response to the second research question, three-factor model can be considered as a more acceptable model in terms of goodness of fit for a better comparative index and higher discriminant validity, that is non-significant correlations between the latent variables (.62, .65 and .49 between peer assessment, feedback and criteria, see Figure 2). Three-factor model did not improve the low factor loadings with the aforementioned eight items either. Only one of the eight items improved in factor loading. Item 37 with a correlation coefficient of .26 under the latent variable of feedforward in five-factor solution gained a correlation of .38 under feedback in five-factor solution.

All in all, CFA revealed a poor fit for FAoW instrument which had been developed based on a Black and Wiliam's (2009) FA model with five factors and a writing model with three stages of pre-writing, writing and post-writing. Except for 13 items under prewriting (Where the learner is going) and the five items of peer assessment, all the other items under two stages of writing and post writing merged due to high correlation. In other words, items showing where the learner is right now functioned the same way that items showing how to get to the objectives.

3.4. The conceptual FAoW model

As the results of the research questions showed earlier, FAoW was not factor structured in the context of this study with the initially developed five constructs. The respondents' experience of FAoW supported three factors of setting criteria, feedback on students' writing tasks and peer-assessment. All the FAoW practices under feed forward and autonomy correlated statistically with the items which measured teachers' feedback on students' current learning (shown as learning evidence in Figure 1). This resulted in assessment in pre- and while-writing process, which is illustrated in a conceptual model in Figure 3.

The model encompasses three main stages of writing assessment and the FA practices that should be implemented on students' writing tasks: Prewriting FA practices, FA practices on the students' current writing task and post writing FA practices which can help students' future improvement and autonomy. As EFL learners' reports showed, the teachers explained learning goals and assessment criteria, encouraged them to brainstorm and develop an outline or mind map. These practices are all tightly related to FA and part of process writing (Hasim, 2014).

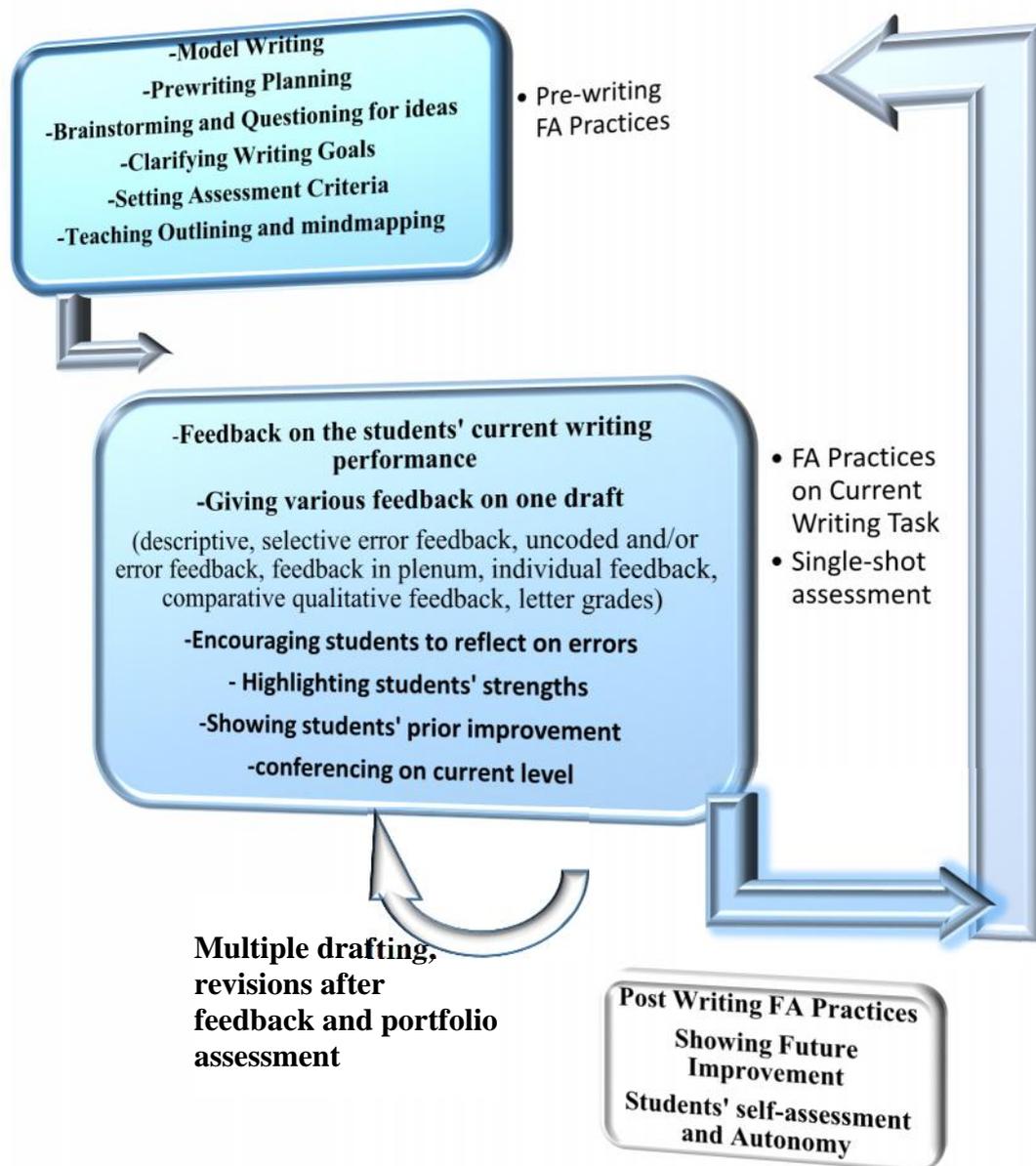


Figure 3. A conceptual model of teachers' practice of FAoW

The existing literature and the findings in this study, however, showed that the practices in the shaded gray parts of FAoW model in Figure 3 were implemented most frequently. In other words, the EFL students in this study did not think they achieved autonomy and independent self-assessment through post-writing FA practices. They learned about the writing goals in pre-writing stage and received single shot assessment on one draft rather than feedback on their revisions through multiple drafting. Hence, with the feedback, which had been usually direct

error correction on a single draft, they moved to the next writing task in the next lesson. It seemed that they were deprived from the teachers' guidelines on how to improve and what to do next for the same task. Similar to studies in other EFL contexts (e.g. Saliu-Abdulahi, 2017), teachers delivered feedback to a finished text instead of asking for resubmission of the text for new assessment.

4. DISCUSSION and CONCLUSION

The specification of FAoW construct through models of FA and writing feedback and its operationalization was the initial stage of instrument development and the aim of an earlier study by the authors. In this study, the instrument was piloted through interview with three EFL learners for qualitative analysis of their comprehension. Subsequently, it was administered in large scale for factor structuring and construct validation through CFA.

The findings of the EFL learners' verbal report confirmed their understanding of the underlying constructs, with the help of examples, definitions and translation notes in the FAoW instrument. In line with Naghdipour (2017), Abdollahzadeh (2010) and Rahimi (2013), the interview findings showed that many of the FAoW practices had never been experienced by EFL learners and that product-based writing and teachers' direct error correction was very common among EFL teachers in writing classrooms. Abdollahzadeh's (2010) study did not aim at the construct of FAoW and only focused on writing strategies among the same population of undergraduate EFL students through large scale questionnaire survey and semi-structured interview. However, the metacognitive strategies in his study overlapped with many of the practices in pre-writing stage in FAoW instrument such as planning for writing, free-writing, awareness of writing purpose and brainstorming. The most common writing strategies among EFL learners were found by him to be metacognitive strategies, FAoW practices known as feed up in this study.

With reference to research questions in this study, our data could not fit in five-factor solution model and the construct of FAoW was found to have a better discriminant validity through three-factor solution. The three-factor model consisted of prewriting (setting assessment criteria) and writing (feedback and peer assessment) and left no post writing stage, which is equally, if not more, crucial in FAoW. The practices in prewriting stage formed criteria (known as feed up). Items under three factors (feedback on current writing, feed forward and autonomy) had to merge for a higher discriminant validity. In the literature, far too little attention has been paid so far to operationalizing the theoretical FA frameworks and writing models by accumulating a comprehensive list of formative feedback practices in writing. Carless (2007) similarly referred to this gap and the existing challenges in implementing the theoretical insights of FA from the literature.

Three-factor structuring was developed for two reasons, firstly due to the strong covariance of the items under feed forward, feedback on current state of writing and autonomy in five-factor structure and secondly because many of the items were theoretically measuring feedback while and after writing. The three highly correlated factors merged into one factor under the name of feedback as feedback was most inclusive of all the practices/ items. It seems that the student respondents in our study perceived the feedback they received on their writing tasks in writing stage as contributors to achieve autonomy and the ability to self-assess.

In addition to statistical justification, modifying the five-factor solution into three factor was theoretically plausible. The items under the three variables dominantly measured teachers' feedback in three stages, before, while and after writing and implementing them for achieving autonomy. Almost all of the items under the three merged factors were directly or indirectly measuring feedback. Furthermore, three stages of writing in Hattie and Timperley's (2007) model of feedback which had initially been used in the development of FAoW instrument could theoretically justify the possibility of merging three variables into one latent variable under the name of feedback and try FAoW model fit with three latent variables.

Abdollahzadeh's (2010) finding of higher frequency of metacognitive strategies (feed up in this study) can also corroborate our findings to the second research question as the practices or items under this construct were distinct from feedback and feed forward in writing and post writing stage. In other words, the students receive feedback on their writing performance in one stage; post-writing stage practices which move learners forward and make them more autonomous through reflection and self-assessment highly correlated with various forms of feedback which is given to students' current writing performance. This was also confirmed through the three participants in the interview. It probably indicated the misconception among EFL learners and maybe their teachers that single stage feedback can promote learners' writing ability to the level of autonomy.

A possible explanation for high correlation between 'current learning evidence' and 'autonomy' is EFL learners' experience of product writing which makes them believe they can progress and write more autonomously through various feedback that they receive in single writing drafts mostly in the form of direct error correction. It seems that their teachers set the criteria for assessment and showed the goals of writing in pre-writing stage; subsequent to the pre-writing stage they implemented all assessment feedback in one stage for students' single writing performance. Apparently, this way of assessment is believed to move learners forward and help them achieve independence and autonomy over time.

FAoW framework with five factors of FA and three stages of writing was not fit for the data collected in this study. Hence, it is probably possible to hypothesize that FA is not utilized in the current EFL context. This can partly be supported with the findings from Naghdipour (2016); Birjandi and Hadidi Tamjid (2012) and Rahimi (2009), who note that writing assessment in Iran follows a product-based tradition and feedback in a single stage. It is characterized by the teachers' focus on students' final products, which is not followed by students' further reflection on the received feedback. Many of the researches in Iranian undergraduate classrooms (e.g. Ketabi, 2015) and in other EFL contexts (e.g. Havnes, Smith, Dysthe & Ludvigsen, 2012; Saliu-Abdulahi, et al., 2017) confirm that assessment is not formative and lacks alternative approaches and various forms of FA.

Construct validation of FAoW instrument could not result in all the five underlying construct being confirmed by EFL learners in this study. Although the instrument was comprehensible for the participants in the qualitative phase of this study and seemed to be a valid measure for identifying students' experience of FAoW, when the factor structuring was analyzed for five factors, the model did not fit the data. For construct validity of FAoW instrument, three-factor solution could reveal a slightly better fit particularly in discriminant validity. The poor model fit of five-factor FAoW in the Iranian EFL context could suggest that the teachers set criteria and show objectives for the writing tasks in pre-writing stage, then incorporate feedback on the students' writing assignment, the feedback which is usually in the form of direct error correction of the form. Feedback is hardly utilized in this context to move learners forward. Feedback on one draft in the context of this study does not feature what Hawe and Parr (2013, p. 215) viewed as an effective practice to promote students' awareness about their improvement. Assessment in the context of this study is on "near-finished products" with the teacher fixing up mistakes not "developmental works in progress".

5. PEDAGOGICAL IMPLICATIONS

The findings of this study provide a set of FAoW practices suggesting an ideal FAoW model, which can be compared with what is actually being employed in EFL contexts. They complement the findings of earlier studies since they show that the practice of writing in EFL classrooms is single drafting and based on assessing the final writing draft. In addition to its theoretical contribution, this study has pedagogical implications for language education contexts. What seems to be missing in writing classrooms is showing the future trend and

helping students how to revise the next drafts by implementing the feedback they have received. Traditional product-based approaches are still the frequent practice and the teachers often offer feedback on different aspects of the students' final draft at one time. The teachers need to encourage further drafting and revision of students' work.

FAoW instrument in this study was validated to identify the teachers' implementation of FAoW practices in the view of EFL learners; although the framework did not fit well with the data in this context, the instrument may have the potential to be utilized by other researchers in other contexts and writing classrooms as it is an operationalized model which can contribute to the utilization of FA. Hence it can be utilized by students, curriculum developers of writing programs and even the teachers (despite its wording) to evaluate the extent to which writing assessment is formative. If its construct is validated in other EFL and international contexts, the developed instrument can be used as a guideline for the teachers as well to know how FA is practiced. The instrument can also be employed by researchers as a classroom observation checklist to measure FAoW practices. The results of this study can additionally raise the awareness of those teachers who are not utilizing FA and are mainly concerned with showing learners their current state of learning rather than the future goals. The developed instrument can additionally pave the way for writing program designers and curriculum developers to implement FA in writing classrooms and utilize various assessment practices prior, while and after the writing stage.

FAoW is a vast area and can include any classroom activity as long as it aims to improve future performance. Multidimensionality of FAoW practices in the instrument was an inevitable problem for the researchers who aimed to develop an instrument with items which needed to tap a single dimension each. The researchers benefitted from both writing feedback model and FA as the theoretical foundation and sought to connect writing with FA. This could probably be assumed as one limitation in this study which could have affected the goodness of fit indices.

The generalizability of FAoW instrument as a measure to reflect teachers' practice of FA in writing classrooms is, therefore, subject to certain limitations. Poor goodness-of-fit statistics in this study makes generalizing the findings to the Iranian EFL teachers' very hard. Overall indices need to be locally justified through further research to provide more specific information about the acceptability and utility of the solution. These limitations made the researchers in this study consider caution when generalizing the findings and try to suffice to the conclusion that the assessment in writing classes in the context under this study seems to be practiced with three rather than five factors, clarifying assessment criteria and writing goals in prewriting stage, peer assessment and feedback in one stage to the final product.

Acknowledgement

This work was supported by NO funding agency.

ORCID

Elaheh Tavakoli  <https://orcid.org/0000-0002-3428-2603>

Mohammad Reza Amirian  <https://orcid.org/0000-0002-3361-731X>

6. REFERENCES

- Arbuckle, J. L. (2012). Amos [Computer software]. Chicago, IL: SPSS.
- Assessment Reform Group. (2002). *Assessment for learning: 10 principles. Research based principles to guide classroom practice*. London, UK: Retrieved from http://languagetesting.info/features/afl/4031afl_principles.pdf
- Abdollahzadeh (2010). Undergraduate Iranian EFL learners' use of writing strategies. *Writing & Pedagogy*, 2(1), 65-90.

- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Birjandi, P., & Hadidi Tamjid, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education*, 37(5), 513-533.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-73.
- Black, P. & Wiliam, D. (2006). *Assessment for learning in the classroom*. In J. Gardner (Ed.), *Assessment and learning*, 9-25. London: Sage.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability* (formerly: *Journal of Personnel Evaluation in Education*), 21(1), 5.
- Brown, T. (2015). *Confirmatory analysis for applied research* (2nd ed.). New York: The Guilford Press.
- Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice*, 8(2), 153-169.
- Burner, T. (2015). Processes of change when using portfolios to enhance formative assessment of writing. *Assessment Matters*, 9(2), 53-79.
- Burner, T. (2016). Formative assessment of writing in English as a foreign language. *Scandinavian Journal of Educational Research*, 60(6), 626-648.
- Elahinia, H. (2004). *Assessment of writing through portfolios and achievement tests*. Unpublished Masters thesis, Teacher Training University, Iran.
- Carless, D. (2007). Learning oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Feng, H. (2007). *Senior ESOL students' experiences of and attitudes towards formative assessment in mainstream secondary classrooms*. Unpublished Masters thesis, University of Canterbury, New Zealand.
- Ghoorchaei, B. Tavakoli, M. & Nejad Ansari, D. (2010). The impact of portfolio assessment on Iranian EFL students' essay writing: A process-oriented approach. *GEMA Online Journal of Language Studies*, 10 (3), 35-51.
- Graham, S., Harris, K.R., Fink, B., & MacArthur, C.A., (2001). Teacher Efficacy in Writing: A Construct Validation with Primary Grade Teachers, *Scientific Studies of Reading*, 5(2), 177-202.
- Hasim, Z. (2014). An integration of a process approach and formative assessment into the development of teaching and learning of ESL writing in a Malaysian university: A sociocultural perspective (Doctoral dissertation, University of Waikato).
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38(1), 21-27.
- Hawe, E., & Parr, J. M. (2013). Assessment for Learning-Form and Substance in Writing Lessons. In *European Conference on Educational Research (ECER) Conference*. Istanbul, Turkey.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.
- Javaherbakhsh, M. R. (2010). The impact of self-assessment on Iranian EFL learners' writing skill. *English Language Teaching*, 3(2), 213-218.
- Johnson, R. C., & Riazi, A. M. (2017). Validation of a locally created and rated writing test used for placement in a higher education EFL program. *Assessing Writing*, 32, 85-104.

- Ketabi, S. (2015). Different methods of assessing writing among EFL teachers in Iran. *International Journal of Research Studies in Language Learning*, 5(2), 3-15.
- Lee, I. (2003). L2 writing teachers' perspectives, practices and problems regarding error feedback. *Assessing Writing*, 8(3), 216-237.
- Lee, I. (2007). Assessment for learning: integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *The Canadian Modern Language Review*, 64(1), 199-213.
- Lee, I. (2011). Formative Assessment in EFL Writing: An Exploratory Case Study, Changing English: *Studies in Culture and Education*, 18(1), 99-111.
- Lee, I. & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34-50.
- Mak, P., & Lee, I. (2014). Implementing assessment for learning in L2 writing: An activity theory perspective. *System*, 47, 73-87.
- Moradan, A., & Hedayati, N. (2011). The impact of portfolios and conferencing on Iranian EFL writing skill. *Journal of English Language Teaching and Learning*, 8, 115-141.
- Mosmery, P., & Barzegar, R. (2015). The effects of using peer, self, and teacher-assessment on Iranian EFL learners' writing ability at three levels of task complexity. *International Journal of Research Studies in Language Learning*, 4(4), 15-27.
- Naghdipour, B. (2016). English writing instruction in Iran: Implications for second language writing curriculum and pedagogy. *Second Language Writing Journal*, 32, 81-87.
- Naghdipour, B. (2017). Incorporating formative assessment in Iranian EFL writing: A case study. *The Curriculum Journal*, 28(2), 283-299.
- Naghdipour, B., & Koç, S. (2015). The evaluation of a teaching intervention in Iranian EFL writing. *The Asia-Pacific Education Researcher*, 24(2), 389-398.
- Nezakatgoo, B. (2005). *The effects of writing and portfolio on final examination scores and mastering mechanics of writing of EFL students*. Unpublished Master thesis, Allame Tabtaba'i University, Tehran, Iran.
- Pallant, J. (2007). SPSS Survival Manual, A Step by Step Guide to Data Analysis using SPSS for Windows, third edition, In *SPSS Survival Manual*. Open University Press, New York.
- Pallant, J., & Manual, S. S. (2007). A step by step guide to data analysis using SPSS for windows. In *SPSS Survival Manual*. Open University Press, New York.
- Rahimi, M. (2009). The role of teacher's corrective feedback in improving Iranian EFL learners' writing accuracy over time: Is learner's mother tongue relevant? *Reading and Writing*, 22(2), 219-243.
- Rahimi, M. (2013). Is training student reviewers worth its while? A study of how training influences the quality of students' feedback and writing. *Language Teaching Research*, 17(1), 67-89.
- Sadeghi, K., & Rahmati, T. (2017). Integrating assessment as, for, and of learning in a large-scale exam preparation course. *Assessing Writing*, 34, 50-61.
- Saliu-Abdulah, D. (2017). Scaffolding writing development: How formative is the feedback?. *Moderna språk*, 111(1), 127-155.
- Saliu-Abdulah, D., Hellekjær, G. O., & Hertzberg, F. (2017). Teachers'(Formative) Feedback Practices in EFL Writing Classes in Norway. *Journal of Response to Writing*, 3(1), 31-55.
- Sharifi, A., & Hassaskhah, J. (2011). The role of portfolios assessment and reflection on process writing. *Asian EFL Journal*, 13(1), 192-229.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Allyn and Bacon.
- Wingate, U. (2010). The impact of formative feedback on the development of academic writing. *Assessment & Evaluation in Higher Education*, 35(5), 519-533.